Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

MARCUS WEBER[†]

# Improved Perron Cluster Analysis

[†]Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Germany.
eMail: weber@zib.de

ZIB-Report 03-04 (April 2003)

# Improved Perron Cluster Analysis

Marcus Weber[1]

[1] Konrad-Zuse-Zentrum Berlin, Takustr. 7, 14195 Berlin, Germany
eMail: weber@zib.de

**Abstract**

The problem of clustering data can often be transformed into the problem of finding a hidden block diagonal structure in a stochastic matrix. Deuflhard et al. [9] have proposed an algorithm that states the number $k$ of clusters and uses the sign structure of $k$ eigenvectors of the stochastic matrix to solve the cluster problem. Recently Weber and Galliat [8] discovered that this system of eigenvectors can easily be transformed into a system of $k$ membership functions or soft characteristic functions describing the clusters. In this article we explain the corresponding cluster algorithm and point out the underlying theory. By means of numerical examples we explain how the grade of membership can be interpreted.

**Keywords:** clustering, fuzzy clustering, stochastic matrix, almost invariant sets.

**MSC:** 62H30, 15A51, 91C20

## 1 Introduction

Cluster analysis is a method for data reduction. Usually one distinguishes between partitioning and hierarchical cluster methods [2]. A well known partitioning cluster method is the $C$-means cluster analysis. In this method high dimensional data is represented by a few prototypes. Each data point is assigned to exactly one prototype. An improvement of this method is called Fuzzy-$C$-Means (FCM) [7]. In this method a grade of membership $0 \leq v_l^{(s)} \leq 1$ for each data point $l = 1, \ldots, m$ to each of the clusters $s = 1, \ldots, k$ is computed.

In this article we want to apply a similar improvement to a partitioning cluster method called Perron Cluster Cluster Analysis (PCCA, or simply Perron Cluster Analysis) [11, 9]. An advantage of PCCA is that the number of clusters can be computed a priori from a spectral analysis of a stochastic matrix $T$ (see below). Deuflhard et al. used the sign structure of the corresponding eigenvectors of $T$ to identify the clusters. The name Perron Cluster Analysis derives from the Perron-Frobenius theorem for nonnegative matrices which Deuflhard applied to the matrix $T$ to get the so called Perron cluster of eigenvalues [12, 11]. This method leads to problems if components of some eigenvectors are approximately 0.

Weber and Galliat [8] discovered that the values of the components of the eigenvectors yield more information about the clusters than the signs of the components. They proposed an algorithm that again computes the grade of membership $v_l^{(s)}$ for each data point $l$ to each of the clusters $s$. This algorithm is faster and easier to understand and implement, and it has no problems if components of some eigenvector are approximately 0 or $T$ is not irreducible.

A sufficient condition for the solvability of the cluster problem with this improved method (PCCA+) is, that after a certain transformation the data will have the shape of a simplex. In this article we will show that this special structure is also necessary for the solvability of the cluster problem. First we will give some examples for possible cluster problems and the corresponding transformation into the problem of finding a hidden block diagonal structure of a stochastic matrix $T$. After a spectral analysis of $T$ we show that the transformation of the cluster problem indeed leads to a data set with simplex structure.

We will describe the algorithm and give some numerical results. The numerical examples will explain how the grade of membership can be interpreted.

# 2 The improved PCCA

**Sources of the data sets.** In this section we will give two examples for possible sources of data. The first will be a dynamical cluster problem, the second a geometrical cluster problem.

The dynamical cluster problem is defined as in [5, 10]: Assume we have a state space $\Omega$ together with a dynamic process represented by a Markov operator. After decomposition of $\Omega$ into $m$ pairwise disjoint subsets and a realization of the Markov operator via Markov chains we get an $m \times m$ transition matrix $A$ counting the number of transitions between the discretization boxes. If the Markov operator is reversible $A$ will be symmetric. We are interested in metastable subsets of $\Omega$, i.e. a clustering of the $m$ discretization boxes in a way that transitions between boxes of different clusters are rare.

The geometrical clustering is a quite similar problem. Galliat [13, 14, 15] uses PCCA to identify clusters in a data set having a measure of similarity: After reducing the data set, for example via self-organizing maps [16], one gets $m$ codebook vectors. The pairwise measure of similarity of these codebook vectors can be written down in a symmetric $m \times m$-matrix $A$. We are interested in a clustering of the $m$ codebook vectors in a way that similarity between vectors of different clusters is small.

In both cases we get a symmetric matrix $A$ and we are interested in the hidden block diagonal structure of $A$.

**Transformation of $A$ into a stochastic matrix.** In the following sections we describe the derivation of the improved PCCA to identify clusters of similar objects or a union of discretization boxes of a state space creating metastable subsets.

The algorithm is designed for symmetric matrices $A$ and for the case that we have a representative for each cluster, that is an object which is only similar to some objects inside the cluster but not to objects outside. The number $m$ of objects is limited by the computability of the $m \times m$-matrix $A$ and the solvability of a symmetric $m \times m$ eigenvalue problem.

We are interested in a decomposition of the index set $\{1,\ldots,m\}$ into subsets $C_i \subset \{1,\ldots,m\}, i = 1,\ldots,k$, in a way that the entries $a_{lj}$ of $A$ for $l \in C_i$ and $j \notin C_i$ are almost 0. In other words: After a permutation of the indexes $A$ has a block diagonal-like structure. We will call this simply a hidden block structure of $A$.

In a first step we transform the nonnegative symmetric matrix $A$ with a diagonal matrix $D^{-1}$ into a stochastic matrix $T = D^{-1}A$, i.e. the sum of the elements of $T$ is 1 for each row. $T$ has the same hidden block structure as $A$. We will investigate the hidden block structure of $T$ via spectral analysis.

**Reformulation of the eigenvalue problem.** Since the condition of a symmetric eigenvalue problem is much better than the condition of a general one, we transform the eigenvalue problem for $T$ into a symmetric form.

$$\begin{aligned} & Tu = \lambda u \\ \Leftrightarrow \quad & (D^{-1/2}AD^{-1/2})D^{1/2}u = \lambda D^{1/2}u. \end{aligned} \tag{1}$$

Instead of solving the eigenvalue problem for $T$ we solve it for the symmetric nonnegative matrix $(D^{-1/2}AD^{-1/2})$ and reweight its eigenvectors with $D^{-1/2}$.

We will now give some results from spectral analysis of $T$ which will be used in the following sections. Since we can transform the eigenvalue problem into a symmetric form, the stochastic $m\times m$-matrix $T$ has a real-valued spectrum with $1 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots \lambda_m \geq -1$. $T$ has got an eigenvalue $\lambda_1 = 1$ which is not necessarily simple. The constant vector $\mathbb{1}$ is a corresponding right eigenvector. For the weighted inner product $< x, y >:= x^T D^{-1} y$ the basis of eigenvectors $u^{(1)},\ldots,u^{(m)}$ can be chosen orthonormal.

From [9] we know that the eigenvectors are almost constant for each index set corresponding to a hidden block of $T$. This is the main tool to identify the clusters.

**Orthogonality condition for the representatives.** In this section we will deduce an orthogonality condition for a set of so called representatives of the clusters. And we use this orthogonality condition to define a grade of membership for each row of $T$ to each of the representative rows. Assume that we have an $m \times m$-matrix $T$ with $k$ hidden blocks, each block $i$ with an index set $C_i \subset \{1,\ldots,m\}$. Further, assume that for each block $i = 1,\ldots,k$ there is an index $\pi(i) \in \{1,\ldots,m\}$ and a row $t_{\pi(i)}$ of $T$ with $t_{\pi(i),j} = 0$ for $j \notin C_i$. We will call such an index $\pi(i)$ a representative for the corresponding block $C_i$. We have the following Kronecker delta relation

$$\mathbb{1}_{C_i}^T t_{\pi(j)} = \delta_{ij}, \quad i,j = 1,\ldots,k \tag{2}$$

for the vector $\mathbb{1}_{C_i}$ which is 1 for components $l \in C_i$ and 0 elsewhere. $\mathbb{1}_{C_i}^T$ is the transposed vector.

The probability $p_{l,C_i}$ for a transition from state $l$ into the index set $C_i$ is given by

$$p_{l,C_i} = \mathbb{1}_{C_i}^T t_l.$$

Hence we get the following orthogonality condition

$$(t_l - \sum_{s=1}^{k} p_{l,C_s} t_{\pi(s)})^T \mathbb{1}_{C_i} = 0, \quad i = 1,\ldots,k, \quad l = 1,\ldots,m. \tag{3}$$

Equation (3) means that an approximation of a row $t_l$ of $T$ via a linear combination of the representative rows $t_{\pi(i)}$ has an error, that is orthogonal to the characteristic functions $1\!\!1_{C_i}$ of the clusters.

Since equation (3) holds for every $i = 1, \ldots, k$ we get for any $a_{ij} \in \mathbb{R}$ and $\bar{u}^{(j)} := \sum_{i=1}^{k} a_{ij} 1\!\!1_{C_i}$

$$(t_l - \sum_{s=1}^{k} p_{l,C_s} t_{\pi(s)})^T \bar{u}^{(j)} = 0, \quad j = 1, \ldots, k, \quad l = 1, \ldots, m. \tag{4}$$

**Definition of the grade of membership.** We can use the orthogonality condition to define a so called grade of membership $v_l^{(i)}$ for row $l$ to the cluster $i$. Since we do not know the index sets $C_i$ a priori, but we know from perturbation theory [9] that the eigenvectors $u^{(i)}$ are almost constant for each $C_i$ (i.e. they are almost equal to the $\bar{u}^{(j)}$ vectors), we can reformulate (4) into the defining equation for $v$ for a given choice $\pi$ of representative rows

$$(t_l - \sum_{s=1}^{k} v_l^{(s)} t_{\pi(s)})^T u^{(j)} = 0, \quad j = 1, \ldots, k, \quad l = 1, \ldots, m. \tag{5}$$

For a dynamical cluster problem the nearly piecewise constant structure of the eigenvectors $u^{(j)}$ is also necessary for high metastability of the corresponding sets [4].

**Positiveness of $v_l^{(s)}$.** In this section we will discuss the conditions for

$$v_l^{(s)} \approx p_{l,C_s} \geq 0. \tag{6}$$

Comparing (4) with (5) we can conclude that the condition (6) is fulfilled if $t_l^T u^{(j)} \approx t_l^T \bar{u}^{(j)}$ for $l = 1, \ldots, m$ and $j = 1, \ldots, k$, or equivalently, if the entries in the rows $t_l$ are small for components where the constant level pattern of $u^{(i)}$ is violated. For dynamical clustering this means that transition states may occur but the system passes through these states very fast.

For the general geometrical cluster problem we know: For any row $t_l$ of $T$ the diagonal element $t_{ll}$ is maximal, because each object is most similar to itself. If the constant level pattern of $u^{(i)}$ is violated in some component $l$, the entry $t_{ll}$ is not small and equation (6) may not be fulfilled. To avoid this situation we use the following trick. After computation of the similarity matrix $A$ we replace its diagonal elements with 0 before computing $T$. The hidden block diagonal structure of $A$ and $T$ has not changed and equation (6) is fulfilled if the objects which are similar to more than one cluster are well separated from other so called transition objects.

**Structure of the data.** In this section we will show that the data we want to cluster has nearly the structure of a simplex. And we will explain how the set of representatives can be found.

Since $u^{(i)}$ is an eigenvector of $T$ the defining equation for $v$ (5) is equivalent to

$$u_l^{(i)} = \sum_{s=1}^{k} v_l^{(s)} u_{\pi(s)}^{(i)}, \quad i = 1, \ldots, k, \quad l = 1, \ldots, m. \tag{7}$$

The constant vector $\mathbb{1}$ is an eigenvector of $T$ corresponding to the eigenvalue $\lambda_1 = 1$. With equation (7) this implies

$$\sum_{s=1}^{k} v_l^{(s)} = 1, \quad l = 1, \ldots, m. \tag{8}$$

From equation (4) we expect that there is a set of representatives such that $v_l^{(i)} \approx p_{l,C_i} \geq 0$. The positiveness of $v$ together with equations (7) and (8) means that the vectors $u_l = (u_l^{(1)}, \ldots, u_l^{(k)})$, for $l = 1, \ldots, m$, are convex combinations of the representatives $u_{\pi(i)}$. In other words the $m$ vectors $u_l \in \mathbb{R}^k$ lie inside a simplex and the $k$ representatives $u_{\pi(i)}$ are its vertices. Once one has found the vertices of the simplex one can compute the grade of membership $v$ via inverting equation (7). Equation (8) together with the positiveness of $v$ makes it possible to interpret $v^{(s)}$ as space covering soft characteristic functions of the corresponding clusters $s$.

**PCCA+ algorithm.** In this section we will explain the improved PCCA algorithm to identify the representatives $\pi(i), i = 1, \ldots, k$, if the number $k$ of clusters is known. This is equivalent to searching for the vertices of a simplex-like data set.

1. Compute the $k$ highest eigenvalues of the $m \times m$-matrix $T$ and the corresponding eigenvectors $u^{(1)}, \ldots, u^{(k)}$ via (1). This is the most time consuming step. The dataset we have to cluster comprises the vectors $u_l := (u_l^{(1)}, \ldots, u_l^{(k)}), l = 1, \ldots, m$.

2. Find two vectors $u_{\pi(1)}$ and $u_{\pi(2)}$ maximizing the distance $||u_k - u_l||, k, l = 1, \ldots, m$ among all data points.

3. For $i = 3, \ldots, k$ find the vector $u_{\pi(i)}$ having the maximal distance to the hyperplane spanned by the vectors $u_{\pi(1)}, \ldots, u_{\pi(i-1)}$.

4. Solve equation (7) to get the grade of membership $v_l^{(i)}$ of the $l^{th}$ element to the $i^{th}$ cluster.

In general the positiveness of $v$ is not fulfilled. The deviation of the shape of the data set from a simplex structure can be fixed by the indicator $\theta$ [8]

$$\theta := \min_{i,l} v_l^{(i)} \leq 0.$$

**Quantification of (6).** The linear equation for $v_l^{(s)}$ (7) can be interpreted as perturbation of the linear equation for $p_{l,C_s}$ (4). Therefore the deviation (6) depends on the relative error of $t_l^T u^{(j)} \approx t_l^T \bar{u}^{(j)}$ and the relative condition number of the matrix $U := (u_{\pi(j)}^{(i)})_{i,j=1,\ldots,k}$, whereas the condition number depends on the quality of the representatives. A well conditioned cluster problem leads to well separated vertices in the PCCA+ algorithm and therefore to a low condition number of $U$ and to a good approximation (6).

**Number of clusters $k$.** In this section we will interpret the vectors $v^{(i)}, i = 1, \dots, k$, as soft characteristic functions of the clusters as in [6, 8]. In [6] we have shown that for a dynamical two cluster problem the soft characteristic functions we get from the above algorithm are optimal in the sence of maximizing metastability of the corresponding clusters.

The vectors $v^{(i)}$ span the same vector space as the eigenvectors $u^{(i)}$ of $T$ because of equation (7). Therefore we have

$$v^{(i)} = \sum_{s=1}^{k} \alpha_{i,s} u^{(s)}$$

with uniquely defined scalars $\alpha_{i,s} \in \mathbb{R}$ and

$$Tv^{(i)} = \sum_{s=1}^{k} \lambda_s \alpha_{i,s} u^{(s)}.$$

Almost invariance of $v^{(i)}$ with regard to $T$ implies that $\lambda_s \approx 1$ for $s = 1, \dots, k$. For a dynamical cluster problem this is a necessary condition for the number of clusters $k$. From perturbation theory we expect a spectral gap between the discrete spectrum of the Markov operator with eigenvalues near 1 and the continuous part of the spectrum bounded away from 1 [3]. This gap should be visible in the spectrum of $T$, too.

In some numerical cases this gap is not visible. Therefore another condition can be useful: For both, the geometrical and the dynamical cluster problem, the data $u_l$ has nearly the shape of a simplex, if the number of clusters $k$ is fixed correctly as we have shown above. In this case the necessary condition for the number of clusters is $\theta \approx 0$. Note that always $\theta = 0$ for $k = 2$ [8].

# 3 Numerical Examples

**Dynamical Data (3-butenal).** In the first example we compute metastable subsets for 3-butenal, see figure 1. We only examine configurational changes of 3-butenal which are indicated by the value of the marked C-C dihedral angle $\varphi$. Via Hybrid Monte Carlo (HMC) one computes 5 Markov chains representing the dynamics of the molecule at a temperature of $300K$ with a characteristic time span of $\tau = 40$fs. The computation was done with the software of Cordes see [6]. The domain of the dihedral angle $\varphi \in [-180°, 180°]$ was decomposed into 90 equidistant intervals, the discretization boxes $m = 90$.

After the simulation one counts the transition number between each pair of discretization boxes. A transition from box $i$ into box $j$ is counted for $A_{ij}$ and $A_{ji}$, too. So we get a symmetric matrix $A$.

Then we apply PCCA+ as described above. Table 1 points out the results for different numbers of clusters $k$.

The eigenvalues $\lambda \leq 0.50$ where interpreted as continuous part of the spectrum of the discretized Markov operator. So the suggested number of clusters is $k = 3$. Since the absolute value of $\theta$ is low for $k = 3$ this is a good choice. And if we plot the transformed data points $u_1, \dots, u_{90}$, see figure 1, we can clearly find the simplex structure.

One can see that the chosen dihedral has three so called conformations. The simplex shows that there are transition states between any pair of conformations.

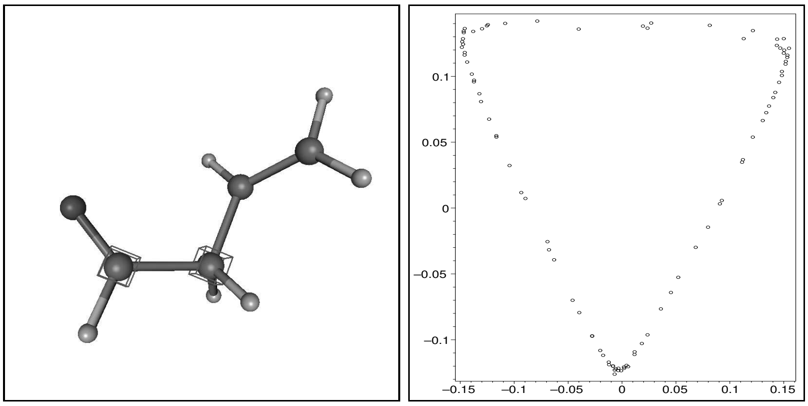


Figure 1: L) One conformation of 3-butenal. Conformational changes correspond to different dihedral angles $\varphi$ of the marked C-C bond. R) The data points $u_l \in \mathbb{R}^3$ for $l = 1, \dots, 90$. Since $u^{(1)}$ is constant, we have left out the first coordinate of $u_l$ in this 2D-projection. In evidence we can see the simplex structure of the data set.

Table 1: Indicator and eigenvalues for different $k$.

| k | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| $\theta$ | 0 | -0.08 | -0.46 | -0.37 | -0.62 | -0.67 | -0.69 | -0.89 | -0.88 |
| $\lambda$ | 0.94 | 0.90 | 0.50 | 0.49 | 0.32 | 0.31 | 0.23 | 0.22 | 0.20 |

**Geometrical Data (2D data set).** In this section we want to give an example for a geometrical clustering. We solve the cluster problem for a $2D$ data set with $m = 30$ data points $d_i, i = 1, \dots, m$, see figure 2. The measure of similarity in this example is an exponential function

$$A_{ij} := \exp(-\mu \, ||d_i - d_j||), \quad i, j = 1, \dots, m, i \neq j,$$

where $A_{ii} := 0$ (see p.4). Another advantage of setting diagonal elements to 0 is, that for higher values of $\mu$ the transition matrix does not converge to unity and therefore not every eigenvalue of $T$ converges to 1. We tested the PCCA+ for different parameters $\mu$ and found out that the higher $\mu$ the closer $\lambda_2$ and $\lambda_3$ move towards 1. Tables 2 and 3 point out the results for $\mu = 2$ and $\mu = 6$ and different numbers of clusters $k$.

Table 2: Indicator and eigenvalues for different $k$. $\mu = 2$

| k | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| $\theta$ | 0 | -0.04 | -0.26 | -0.26 | -0.14 | -0.92 | -0.92 | -0.92 | -0.96 |
| $\lambda$ | 0.86 | 0.63 | 0.14 | -0.11 | -0.12 | -0.12 | -0.13 | -0.13 | -0.13 |

Since $\theta \approx 0$ for $k = 3$, this is the right number of clusters. Figure 2R shows the corresponding simplex for $\mu = 2$ and $k = 3$. If we compute the grade of membership $v$, then we can assign each index $i$ to the cluster with the highest grade of membership $\max_s v_i^{(s)}$. The result of this cluster method for $\mu = 2$ and

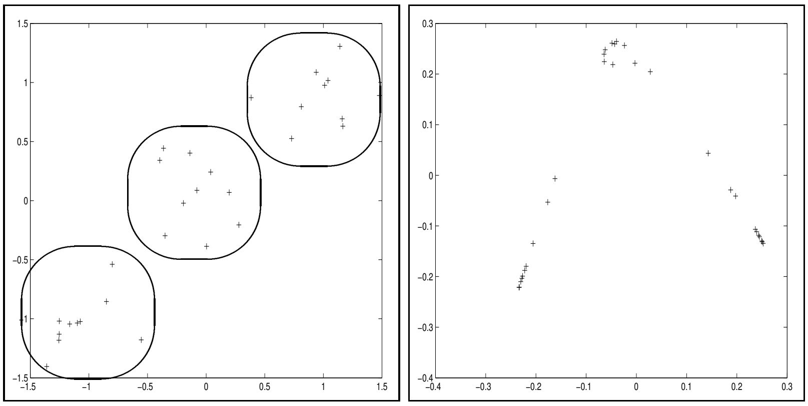

Figure 2: L) 30 data points for a geometrical clustering with $k = 3$. R) Simplex for the geometrical cluster problem: The data points $u_l \in \mathbb{R}^3$ for $l = 1, \ldots, 30$. Since $u^{(1)}$ is constant, we have left out the first coordinate of $u_l$ in this 2D-projection. We can see the simplex structure of the data set. $\mu = 2$

Table 3: Indicator and eigenvalues for different $k$. $\mu = 6$

| k | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| $\theta$ | 0 | -0.004 | -0.04 | -0.04 | -0.20 | -0.18 | -0.24 | -0.24 | -0.24 |
| $\lambda$ | 0.99 | 0.97 | 0.55 | 0.48 | 0.34 | 0.27 | 0.20 | -0.28 | -0.28 |

$k = 3$ is shown in figure 2L. For $\mu = 6$ we get the same result.

**Comparison with FCM.** We have clustered the example from figure 2 with the Fuzzy-C-Means algorithm [1]. If we again assign each data point to the cluster with the highest grade of membership the result is the same as in the PCCA+ method. The main differences between FCM and PCCA+ are:

1) FCM generates prototypes which are located in the centers of their clusters. PCCA+ computes representatives among the data points with maximal dissimilarity.

2) PCCA+ computes the number of clusters a priori.

# 4 Conclusion

The PCCA algorithm computes the number of clusters for dynamical cluster problems via spectral analysis of a stochastic matrix. In this article we improved PCCA by means of linear algebra.

PCCA+ is a useful and simple algorithm for many small sized cluster problems that can be transformed into the problem of finding a hidden block diagonal structure in a stochastic matrix $T$.

This article has shown that under certain conditions, existence of representatives and well behaving transition objects, the transformation leads to a data set which has the shape of a simplex.

The vertices of the simplex are the representatives and the location of the data points in the interior of the simplex determines their grade of membership to each of the representatives.

Therefore PCCA+ can also be used to characterize transition states in dynamical cluster problems.

**Acknowledgements.** I want to thank Peter Deuflhard and Frank Cordes for many helpful discussions and Daniel Baum for useful hints for this article.

# References

[1] B. Allmendinger. C++ classes for the fuzzy-c-means algorithm (version 1.0). http://www.neurocomputing.de, February 2000.

[2] B.D.Ripley. *Pattern Recognition and Neural Networks.* Camebridge University Press, 1996.

[3] Ch.Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm and Application to Biomolecules.* Habilitation Thesis, Dept. of mathematics and computer science, Free University Berlin, 1998.

[4] Ch.Schütte and W.Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. *Accepted for Handbook of Numerical Analysis, special volume on computational chemistry*, 2002.

[5] Ch.Schütte, A.Fischer, W.Huisinga and P.Deuflhard. A direct approach to conformational dynamics based on hybrid monte carlo. *J. Comput. Phys., Special Issue on Computational Biophysics*, 151:pp. 146–168, 1999.

[6] F.Cordes, M.Weber and J.Schmidt-Ehrenberg. Metastable conformations via successive perron cluster cluster analysis of dihedrals. *Report 02-40, ZIB*, 2002.

[7] J.C.Bezdek and S.K.Pal. *Fuzzy Models for Pattern Recognition.* IEEE Press, New York, 1992.

[8] M.Weber and T.Galliat. Characterization of transition states in conformational dynamics using fuzzy sets. *Report 02-12, ZIB*, 2002.

[9] P. Deuflhard, W.Huisinga, A.Fischer and Ch.Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled markov chains. *Lin.Alg.Appl.*, 315:pp. 39–59, 2000.

[10] P.Deuflhard. From molecular dynamics to conformational dynamics in drug design. In *Trends in Nonlinear Analysis.* M. Kirkilionis, S. Krömker, R. Rannacher, F. Tomi, editors., Springer, 2003.

[11] P.Deuflhard and A.Hohmann. *Numerische Mathematik I.* Walter de Gruyter Berlin, New York, $3^{rd}$ edition, 2002.

[12] R.B.Bapat and T.E.S.Raghavan. *Nonnegative Matrices and Applications.* Cambridge University Press, 1997.

[13] T.Galliat. *Adaptive Multilevel Cluster Analysis by Self-Organizing Box Maps.* PhD thesis, Department of Mathematics and Computer Science, Free University of Berlin, March 2002.

[14] T.Galliat and P.Deuflhard. Adaptive hierarchical cluster analysis by self-organizing box maps. *Report 00-13, ZIB*, 2000.

[15] T.Galliat,W.Huisinga and P.Deuflhard. Self-organizing maps combined with eigenmode analysis for automated cluster identification. *Proc. of the 2nd Intern. ICSC Symposium on Neural Computation, ISCS Academic Press*, 2000.

[16] T.Kohonen. *Self-Organizing Maps.* Springer, Berlin, Heidelberg, New York, 3 edition, 2001.